

CSC 4740/6740 Data Mining

Assignment 3

Due Date: 11:59 am, October 31, 2022

1. (100 points) Implementing PageRank algorithm.

Example 1: The toy dataset is the following graph. The PageRank values are already known. We can use it to check your program.

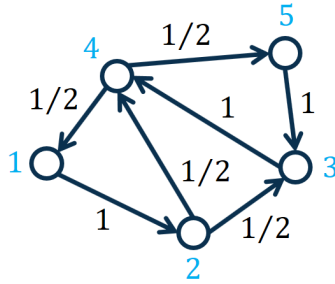


Figure 1: A toy graph for computing PageRank. The number on the edge represents the transition probability from one node to another.

The PageRank values are given in the following table (given that the decay factor $c = 0.85$):

Nodes	PageRank Values
1	0.1556
2	0.1622
3	0.2312
4	0.2955
5	0.1556

PageRank:

Compute the PageRank value of each node in the graph. Please refer to the slides for more details about the PageRank method. The key PageRank equation is as follows.

$$\mathbf{r} = c\mathbf{P}^T\mathbf{r} + (1 - c)\mathbf{1}/n$$

where \mathbf{r} represents the $n \times 1$ PageRank vector with each element r_i representing the PageRank value of node i , n represents the number of nodes in the graph, \mathbf{P} represents the $n \times n$ transition probability matrix with each element $\mathbf{P}_{i,j} = p_{i,j} = \frac{1}{d_i}$ representing the transition probability from node i to node j , d_i represents the degree of node i , \mathbf{P}^T represents the transpose of \mathbf{P} , $c \in (0,1)$ represents a decay factor, $\mathbf{1}$ represents a $n \times 1$ vector of all 1's, and n represents the number of nodes in the graph.

Please see the slides for more details.

In this assignment, we set the decay factor $c = 0.85$ and set the number of iterations to 30.

Implementation: Please implement the Power Iteration algorithm for computing PageRank in either Matlab or Python.

You need to submit the complete Matlab or Python file into iCollege. You are required to work on only one of them.

You need to copy and paste the outputs of your algorithm in the console/terminal for the Example 1 above.

Output in your Matlab algorithm:

Or the output in your Python algorithm: